FEBS Letters 584 (2010) 1047–1053

FEBS Letters
journal homepage: www.FEBSLetters.org

Vertebrate β-thymosins: Conserved synteny reveals the relationship between those of bony fish and of land vertebrates

John Edwards

Faculty of Biomedical and Life Sciences, University of Glasgow, United Kingdom

ARTICLE INFO

Article history:
Received 18 December 2009
Revised 1 February 2010
Accepted 2 February 2010
Available online 9 February 2010

Edited by Takashi Gojobori

Keywords:
β-Thymosin
Conservation of synteny
Teleost
Tetrapod
Whole genome duplication

ABSTRACT

Using conservation of synteny I show how the four thymosins expressed by teleost fish are related to the three of tetrapods, which is not evident from their protein sequences. This clarification was aided by identification of a novel thymosin of reptilians that replaces the β10 thymosin of mammals. Recent reconstruction of the ancestral vertebrate genome suggests that divergence of β-thymosins began with duplication preceding the two rounds of whole genome duplication.

1. Introduction

β-Thymosins are peptides of 40–45 amino-acid residues [1]. First identified from a cytokine-like activity [2], the archetypical β-thymosin, β4 was later found to be an intracellular protein that sequesters G-actin [3] and thereby influences microfilament dynamics.

β4 has therapeutic potential: as an anti-inflammatory agent [4] and a promoter of wound healing in humans [5] and in relation to acute myocardial damage in a mouse model [6]. β-Thymosins 4, 10 and 15 have shown increased expression associated with malignancy [7–9], and the human β15 may provide a valuable diagnostic for malignancy of prostate cancer [10].

Although it is accepted that mammals express three β-thymosins [11], relationships between these and the numerous β-thymosins of bony fish have not been documented, and a wider vertebrate perspective has been lacking. For example, it has not hitherto been evident which teleost thymosins are the closest relatives of β-thymosins 4 and 15. I show here that groups of teleost β-thymosin genes can be traced back to genes common with the better-known groups of mammalian β-thymosins in a common ancestor of these two lineages. By definition this establishes their relationship as orthologues.

2. Methods

β-Thymosin genes were identified from 5 teleost and 18 tetrapod genomes available in 2008 via the UCSC browser [12], using BLAT searches [13] for translated sequences similar to known β-thymosins. Alignments extending beyond exon 2 were found to be intronless pseudogenes and were excluded. Transcripts encoding proteins similar to known β-thymosins were identified from mRNA and Expressed sequence tag (EST) databases at NCBI using tBLASTn limited to single species, and their coding sequences verified by alignment against respective genomes. For representative transcripts and assembly dates, see Supplementary Figs. 1 and 2.

Preliminary investigation, including principal component analysis (not shown) indicated that relationships between groups of tetrapod and teleost β-thymosins cannot be deduced from their amino-acid sequences. I therefore made use of the observation that genome segments containing stable sets of multiple genes can be traced through evolutionary history between even relatively divergent species of vertebrates. Such stable context of surrounding genes (conserved local synteny) provides a powerful means of establishing orthology.

To document conservation of synteny, a set of 20 human protein sequences was compiled (from the UCSC browser human proteins

Abbreviations: CVL, conserved vertebrate linkage; EST, expressed sequence tag; Mbp, million base pairs; Myr, million years; UTR, untranslated region; WGD, whole genome duplication

E-mail address: john@mblab.gla.ac.uk

0014-5793/$36.00
doi:10.1016/j.febslet.2010.02.004

track) which aligned via tBLASTn with 10 genes in each direction from specified β-thymosins. (For an example, see Supplementary Fig. 3.) These sets were used as inter-specific probes between teleosts by counting BLAT alignments in windows spanning from ±1 to −1 million base pairs (Mbp) of target β-thymosins. To relate β-thymosins of teleosts to those of tetrapods, BLAT alignments (>/=90% identity and >/=90% length) of these flanking homologues were located in the human genome and thus assigned to 7 of the 100+ conserved vertebrate linkage (CVL) blocks described by Nakatani et al. [14]. For an example see Supplementary Fig. 3.

For 3′UTR conservation, unaligned 3′UTRs from transcripts of β4 thymosins of 9 species were submitted to MEME [15], seeking an optimal motif between 5 and 50 bp. A 48-mer sequence was identified, with bases 14–48 virtually identical across all tetrapods. The position-specific scoring matrix for this motif was submitted to MAST [16] to search a database containing 3′UTRs of all β-thymosins of the 9 training species, plus those of all the gene-mapped fish.

Protein sequences were translated from mRNA or ESTs. In the absence of transcripts, exon 3 sequences were identified by their consensus: IEQ[Q/E][K/R][X]$_n$ (Stop), $n < 9$.

Alignments and phylogenetic trees were constructed using Phylogeny.fr [17] with default settings (briefly: MUSCLE alignment, gap removal, and tree construction using maximum likelihood). Sequence alignments were edited and Taylor-coloured [18] in Jalview [19].

3. Results

Orthology between teleost and tetrapod β-thymosins is not apparent from their protein sequences. I therefore first assigned β-thymosins to groups within each of these two lineages, and then used conservation of synteny to relate the two sets of groups. Groups were numbered retrospectively to correspond.

3.1. Teleost genes

Five β-thymosin genes were identified in zebrafish, four each in the other fish. Orthology for four per species (three only in zebrafish) was readily deduced from conservation of synteny (Fig. 1). Groups 1a and 1b are very similar in protein sequence. In medaka and tetraodon the 1a/1b pairs are located on sister chromosomes (1 and 21, 2 and 3, respectively) of the teleost-specific R3 whole genome duplication (WGD) [20], so 1a/1b pairs presumably arose then. The zebrafish genome lacks a 1b β-thymosin, consistent with its known loss of the R3 WGD sibling of chromosome 9 [21], which bears the 1a copy. Two extra zebrafish thymosins (on chromosomes 1 and 21) lack syntenic equivalents in the other fish (Fig. 1). In all five fish, groups 2 and 3 genes are located close to each other on the same chromosome, with separations ranging from 0.8 Mbp (tetraodon) to 32 Mbp (zebrafish).

3.2. Teleost protein sequences

Despite the few informative positions, phylogenetic trees group teleost β-thymosins largely in agreement with the gene mapping, supporting the three main groups (Fig. 2). Exceptions, misplaced, are zebrafish group 1a and the zebrafish thymosin on chromosome 1. For the former, a motif (of unknown function) in its 3′UTR, previously found remarkably conserved among the β4 (tetrapod) thymosins [22], independently supports its gene-mapped assignment to group 1 (Fig. 3). For the latter, see: relationship of teleost and tetrapod groups and Fig. 6.

Protein sequence similarity allows assignment of β-thymosins of many other fish to the three groups (see Supplementary Fig. 1).

Stickleback flanks	1a	1b	2	3
Stickleback				
1a	20	2	0	1
1b	2	20	0	2
2	0	0	20	2
3	5	3	1	20
Fugu				
1a	19	2	0	2
1b	2	11	1	2
2	0	2	13	1
3	3	4	1	11
Medaka				
1a	18	2	0	1
1b	2	11	0	2
2	0	0	13	1
3	4	3	1	21
Tetraodon				
1a	18	4	0	1
1b	2	15	0	2
2	0	0	19	2
3	4	3	1	20
Zebrafish				
1a	11	2	0	1
2	0	0	10	1
3	2	2	1	7
Chr1	1	0	0	0
Chr21	0	2	0	2

Fig. 1. Conservation of synteny of β-thymosins of five teleost fish. The local context of genes surrounding each β-thymosin of the stickleback was compared in turn with that of each β-thymosin of the other teleost species. BLAT alignments from a set of 20 translated sequences of human homologues of the neighbouring genes were counted in a 2 Mb window. Because blocks of genes around each β-thymosin are much more strongly conserved between species than between different thymosins of a single species, most teleost thymosins could be assigned unequivocally to syntenic, and by inference, orthologous, groups. The same grouping is obtained by probing with flanking sequences of medaka (not shown).

3.3. Tetrapods

The β-thymosins of land vertebrates also fall into three main groups, containing, respectively, the mammalian β4, β10 and β15 thymosins [11], which is confirmed here in Fig. 4. (An earlier report of a seven-member multi-gene family for β4 in humans [23] may have been influenced by inclusion of pseudogenes.) However, instead of β-thymosins similar to mammalian β10, birds and lizards express a distinct set. With unique residue 6 = phenylalanine ("6F"), and expression in zebrafinch and lizard supported by numerous ESTs, these are newly identified here. Although closest to β4 in protein sequence, conservation of synteny shows them to be orthologues of fish group 2, not 1 (see Fig. 5). In addition to the threefold set, searches also identified primate- and rodent-specific duplications of β15 genes, some previously documented [11] [24], and Y chromosomal paralogues of β4 in human (as in [25]) and chimpanzee.

3.4. Relationship of teleost and tetrapod groups

The recent reconstruction of vertebrate ancestral chromosomes, based on human and medaka genomes [14], provides a powerful means of tracing conserved synteny and constructing a hypothesis of the ancestry of the β-thymosin genes. Thus teleost group 1a is orthologous with tetrapod β4, group 2 with "6F" of birds and lizard,

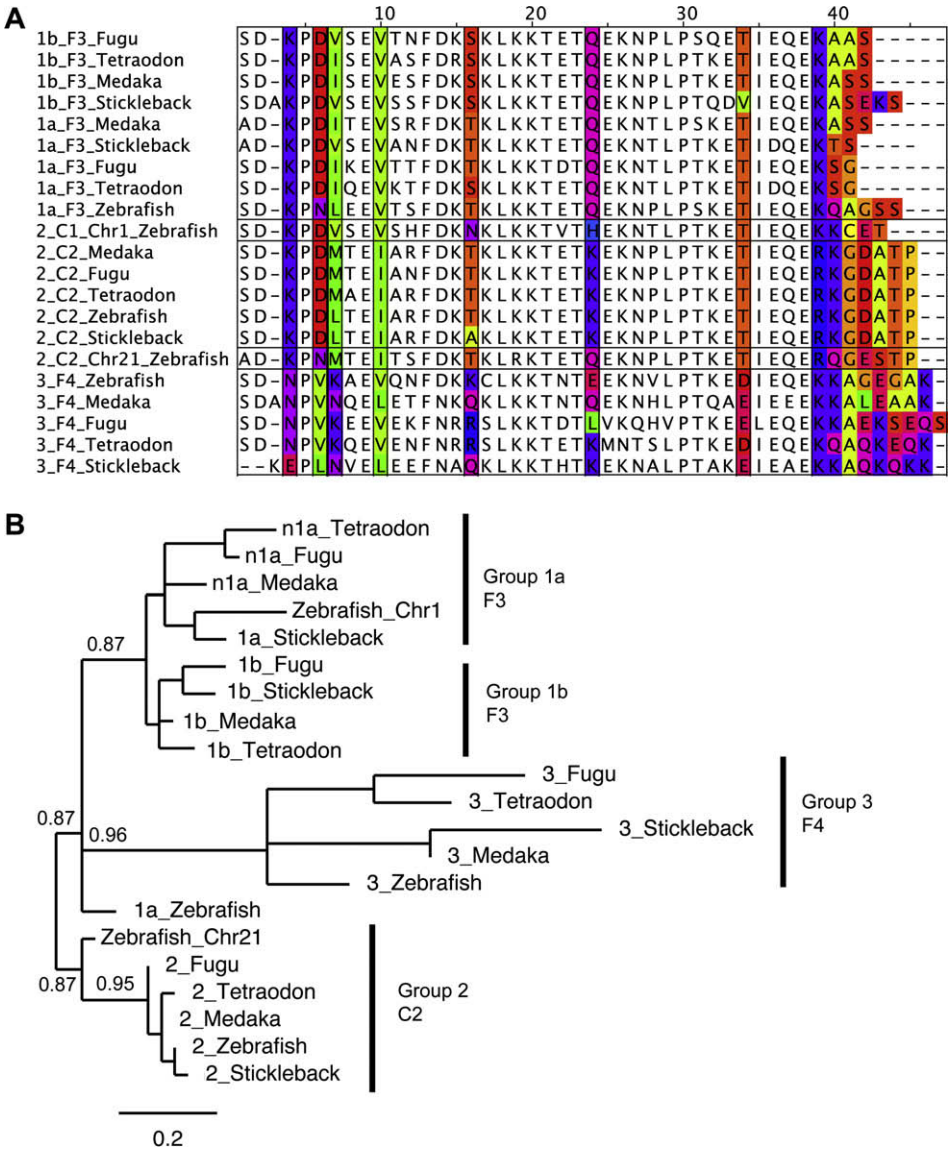

Fig. 2. Teleost thymosin protein sequences. (A) Alignment: Sequences, preceded by group numbers assigned by synteny, are in alignment order. C1, C2, F3 and F4 are conserved vertebrate linkage blocks (see text). Residues are Taylor-coloured where there is a well-conserved difference in character between one group and either of the other two. The gap or residue following position D2 causes a +1 shift relative to usual thymosin numbering. (B) Phylogenetic tree: The tree is rooted between group 1-with-3 and group 2 on the basis of the proposed ancestry (see Fig. 6). Branch support values are the result of an Approximate Likelihood-Ratio Test [17]. Branches with support values < 0.5 are collapsed. Branch lengths are proportional to the number of substitutions per site, indicated by the scale bar. Clustering of zebrafish thymosin 1a and the zebrafish thymosin gene on chromosome 1 are contrary to firm evidence from conservation of synteny.

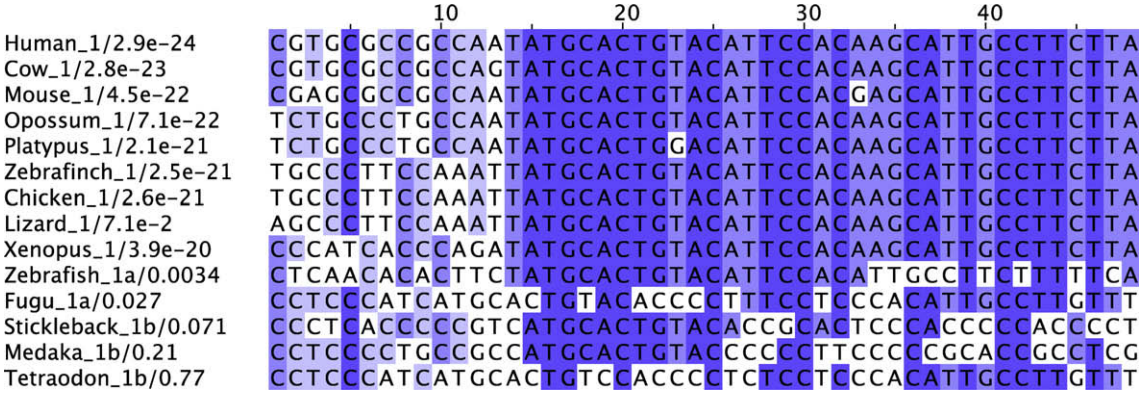

Fig. 3. Conservation of 3′UTR sequence between group 1 thymosins of tetrapods and fish. *E* values are listed after species and group, and all sequences with *E* < 1 are shown. Shading by percentage identity (85, 60, 50).

and group 3 with tetrapod β15, as shown by the continuous columns of flanking sequence alignments to the human genome in Figs. 5 and 6. The assignment to CVLs indicates that these β-thymosins originated from copies on proto-chromosomes F3, C2 and F4, respectively, of the gnathostome common ancestor. Assignment to CVLs also solves the otherwise obscure provenance of mammalian β10 thymosins, which show no conservation of synteny with any other vertebrate thymosins. Alignments of teleost group 2 flanking sequences cluster on human chromosome 5, itself a locus devoid of thymosin, in a CVL block assigned to ancestral proto-chromosome C2. The human β10 locus, on chromosome 2, however, is in CVL block C3, an R1/R2 WGD sibling of C2. (Shown by the discontinuous group 2 columns in Figs. 5 and 6.)

Mapping to CVLs also suggests possible ancestry of the zebrafish β-thymosins on chromosomes 1 and 21, and the *Xenopus* β10-like thymosin, which map to ancestral chromosomes C1, C2

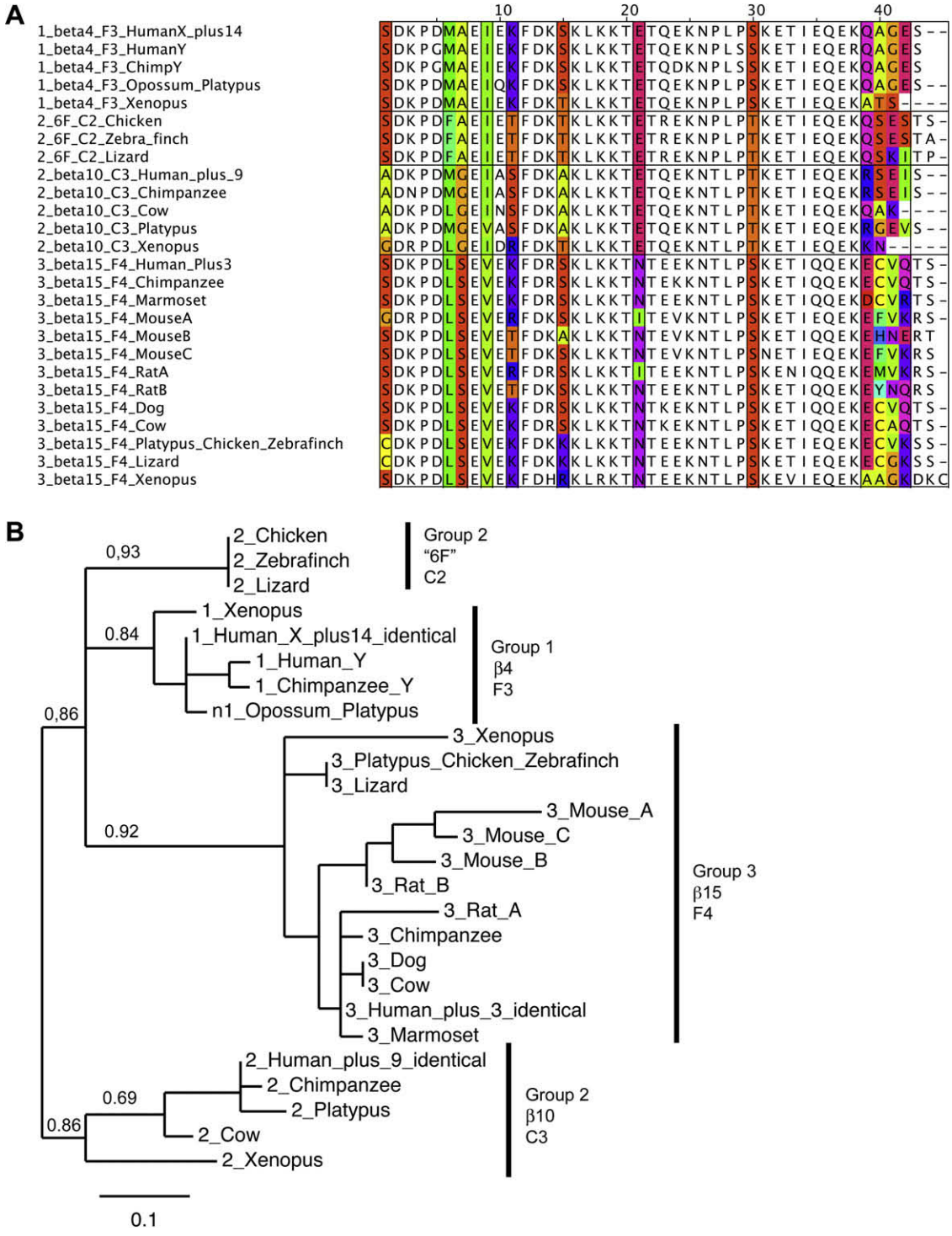

Fig. 4. Tetrapod thymosin protein sequences. (A) Alignment: Details as for Fig. 2A. The chicken, zebra finch and lizard group 2 sequences ("6F") are assigned to group 2 by conserved synteny. Duplications of β15 are suffixed A, B or C in order of chromosomal co-ordinates. (B) Phylogenetic tree: Sets of identical sequences are represented by one. Other details as Fig. 2B.

and C3, respectively (Fig. 6). The zebrafish chromosome 1 gene thus originates from another WGD sibling of ancestral chromosome C, a copy not retained by other vertebrates. That on chromosome 21 (TYB_DANRE in the Swiss Protein database) maps to the same ancestral chromosome as group 2 teleost thymosins, which it closely resembles in protein sequence, and may therefore be an R3 WGD copy lost from the other fish or a late duplication specific to the Cypriniform lineage. The related fish *Misgurnus anguillicaudatus* expresses a protein very similar to TYB_DANRE (see Supplementary Fig. 1). The *Xenopus* β10-like protein is most likely the orthologue of mammalian β10 as its protein sequence suggests (but see Supplementary note 1).

4. Discussion

Assignment of teleost and tetrapod thymosins to proto-chromosomes of their gnathostome common ancestor leads to the hypothesis of their ancestry summarised in Fig. 7. In this the first β-thymosin divergence formed a pair on different vertebrate ancestral proto-chromosomes (C and F) consistent with a very early duplication (>600 million years (Myrs), preceding the R1/R2 WGD). Divergence of group 1/β4 from group 3/β15 was later, following R1 or R2 WGDs. The proximity on the same chromosome of the groups 2 and 3 genes in teleosts must result from chromosome fusion, frequent in the teleost lineage around 400–500 Myrs [14], rather than tandem duplication.

The distinction between reptilian "6F" and mammalian β10 thymosins is a result of these lineages having retained different R1/R2 WGD copies of a gene long diverged from β4. More recently, the teleost-specific R3 WGD formed the 1a/1b pair of teleosts.

Within teleosts and within tetrapods, groups of β-thymosins can be distinguished by comparing their protein sequences (Figs. 2 and 4). Thus differentiation between groups has been retained for some 400 Myrs. However, its particulars correlate little between the two classes. For example, group 3 teleost β thymosins all lack the charge pair 3K4P5D, but tetrapod group 3/β15 retain it. (It is noteworthy, however, that within both taxa, group 3 sequences are more diverse than the other two.) Differentiation

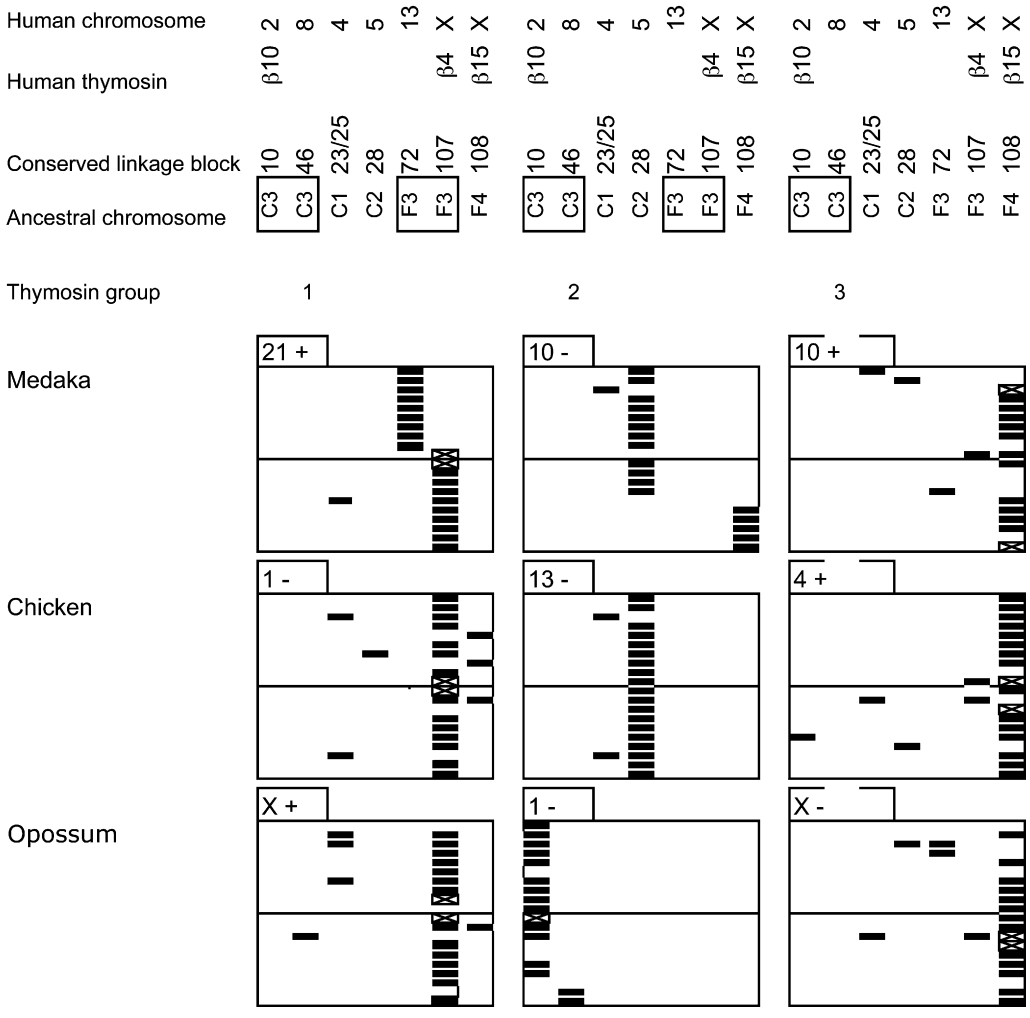

Fig. 5. Assignment of β-thymosin flanking sequences to conserved vertebrate linkage (CVL) blocks. Nakatani et al. [14] identified blocks of genes (CVL blocks) conserved between medaka, chicken and human genomes and provided a detailed map of these on the human genome. Here BLAT alignments of protein sequences translated from human homologues of genes surrounding β-thymosins were located in the human genome, allowing assignment of each β-thymosin gene to a specific CVL block. An example showing details of one set of surrounding genes and its mapping is shown in Supplementary Fig. 3. Each of the nine large boxes displays data from surrounding genes of one β-thymosin of the named probe species and is labelled above left with its chromosome and strand polarity. Within each box is a matrix of, *Y*-axis: the 20 human homologues of genes surrounding the probe thymosin, and *X*-axis: a set of 7 CVL blocks defined by the position of each on a human chromosome [14]. Cells are filled when BLAT finds an alignment of >/=90% identity, and >/=90 length. Thus positions of aligned sequences relative to the probe thymosin are shown, but not their order on the human chromosome, within the CVL. However, where there is conserved synteny with a human thymosin, the two immediately adjacent flanking sequences are indicated with white crosses. CVLs shown are those with more than 4 hits from any probe species. For an example of input data, see Supplementary Fig. 3.

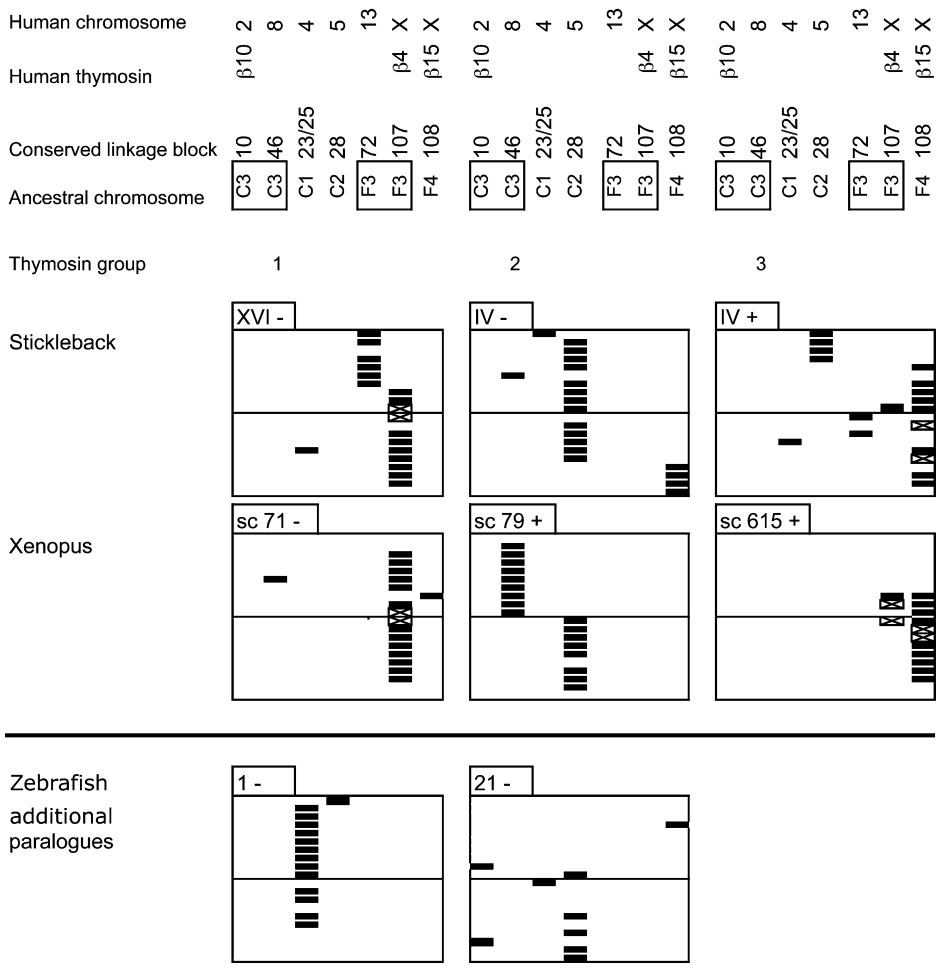

Fig. 6. Assignment of β-thymosin flanking sequences to conserved vertebrate linkage (CVL) blocks. As Fig. 5, except that the strand polarity of plotting stickleback group 1 thymosin flanks is reversed to correspond with medaka in Fig. 5. Upstream sequences of the *Xenopus* β15 homologue were limited to four only by proximity to the end of the scaffold.

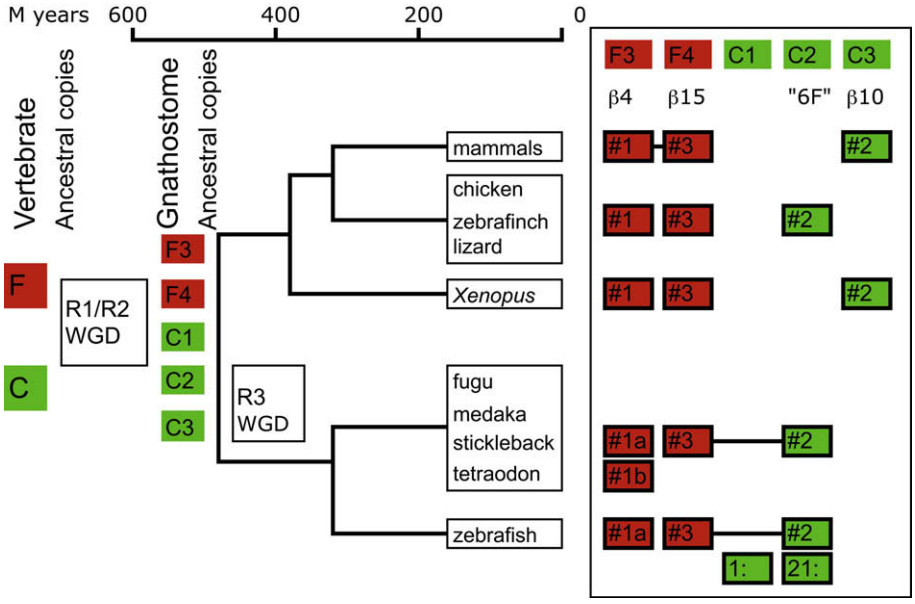

Fig. 7. Proposed ancestry of vertebrate β-thymosin genes. Boxed rectangles: The present day β-thymosins. Unrelated linkage of mammalians β4 and β15 (on X chromosome) and teleost groups 2 and 3 thymosins are indicated by the horizontal joining lines. Unboxed rectangles: Inferred ancestral genes on ancestral proto-chromosomes.

between the β-thymosins may represent fine-tuning of intracellular actin-binding, as explored by swapping experiments between C-termini of mammalian β4 and β15 [26]. It could also relate to other "moonlighting" [5] functions. For example, the methionine replacing leucine at residue 6 is susceptible to oxidation to the sulphoxide, which could weaken interaction extracellularly with actin [27], or signal other interactions [4]. Further explanation would be needed for why differentiation followed distinct paths above and below water.

The orthology described here and the proposed evolutionary history will be valuable for studies of the expression, evolution and function of these small, often highly expressed proteins. It could also provide the basis for a nomenclature applicable across all vertebrates.

Acknowledgments

I thank numerous genome sequencing and assembly groups for sequence assemblies. These include: Bovine Genome Project at Baylor College of Medicine, Broad Institute, Genome Sequencing Center of Washington University in St. Louis, Genoscope, Japanese National Institute of Genetics and University of Tokyo, National Center for Biotechnology Information, US DOE Joint Genome Institute, Wellcome Trust Sanger Institute.

I thank Robin Stevenson, Tony Lawrence and John Young for provoking my interest in β-thymosins and geneticist colleagues, particularly Mark Bailey for critical comments.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2010.02.004.

References

[1] Huff, T., Müller, C.S.G., Otto, A.M., Netzker, R. and Hannappel, E. (2001) ß-Thymosins: small acidic peptides with multiple functions. Int. J. Biochem. Cell Biol. 33, 205–220.
[2] Low, T.L.K., Hu, S.K. and Goldstein, A.L. (1981) Complete amino acid sequence of bovine thymosin β4: a thymic hormone that induces terminal deoxynucleotidyl transferase activity in thymocyte populations. Proc. Natl. Acad. Sci. USA 78, 1166–1170.
[3] Safer, D., Golla, R. and Nachmias, V.T. (1990) Isolation of a 5-kDa actin-sequestering peptide from human blood platelets. Proc. Natl. Acad. Sci. USA 87, 2536–2540.
[4] Young, J.D., Lawrence, A.J., MacLean, A.G., Leung, B.P., McInnes, I.B., Canas, B., Pappin, D.J. and Stevenson, R.D. (1999) Thymosin beta 4 sulfoxide is an anti-inflammatory agent generated by monocytes in the presence of glucocorticoids. Nat. Med. 5, 1424–1427.
[5] Goldstein, A.L., Hannapel, E. and Kleinman, H.K. (2005) Thymosin beta4: actin-sequestering protein moonlights to repair injured tissues. Trends Mol. Med. 11, 421–429.
[6] Srivastava, D., Saxena, A., Dimaio, J. and Bock-Marquette, I. (2007) Thymosin beta 4 is cardioprotective after myocardial infarction. Ann. N.Y. Acad. Sci. 1112, 161–170.
[7] Kobayashi, T., Okada, F., Fujii, N., Tomita, N., Ito, S., Tazawa, H., Aoyama, T., Choi, S.K., Shibata, T., Fujita, H. and Hosokawa, M. (2002) Thymosin-β4 regulates motility and metastasis of malignant mouse fibrosarcoma cells. Am. J. Pathol. 160, 869–882.
[8] Santelli, G., Califano, D., Chiappetta, G., Vento, M.T., Bartoli, P.C., Zullo, F., Trapasso, F., Viglietto, G. and Fusco, A. (1999) Thymosin beta-10 gene over-expression is a general event in human carcinogenesis. Am. J. Pathol. 155, 799–804.
[9] Bao, L., Loda, M., Janmey, P.A., Stewart, R., Anand-Apte, B. and Zetter, B.R. (1996) Thymosin beta15: a novel regulator of tumor motility up-regulated in metastatic prostate cancer. Nat. Med. 2, 1322–1328.
[10] Hutchinson, L.M., Chang, E.L., Becker, C.M., Shih, M.C, Brice, M., DeWolf, W.C., Gaston, S.M. and Zetter, B.R. (2005) Use of thymosin beta 15 as a urinary biomarker in human prostate cancer. Prostate 64, 116–127.
[11] Banyard, J., Hutchinson, L.M. and Zetter, B.R. (2007) Thymosin β-NB is the Human Isoform of Rat Thymosin β15. Ann. NY Acad. Sci. 1112, 286–296.
[12] Karolchik, D. et al. (2008) The UCSC Genome Browser Database: 2008 update. Nucleic Acids Res. 36, D773–779.
[13] Kent, W.J. (2002) BLAT – the BLAST-like alignment tool. Genome Res. 12, 656–664.
[14] Nakatani, Y., Takeda, H., Kohara, Y. and Morishita, S. (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res. 17, 1254–1265.
[15] Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymersProceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28–36, AAAI Press, Menlo Park, CA.
[16] Bailey, T.L. and Gribskov, M. (1998) Combining evidence using *p*-values: application to sequence homology searches. Bioinformatics 14, 48–54.
[17] Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M. and Gascuel, O. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36, W465–W469.
[18] Taylor, W.R. (1997) Residual colours: a proposal for aminochromography. Protein Eng. 10, 743–746.
[19] Clamp, M., Cuff, J., Searle, S.M. and Barton, G.J. (2004) The Jalview Java Alignment Editor. Bioinformatics 20, 426–427.
[20] Meyer, A. and Van de Peer, Y. (2005) From 2R to 3R: evidence for a fish-specific genome duplication. Bioessays 27, 937–945.
[21] Kasahara, M. et al. (2007) The medaka draft genome and insights into vertebrate genome evolution. Nature 447, 714–719.
[22] Duret, L. and Bucher, P. (1997) Searching for regulatory elements in human noncoding sequences. Curr. Opin. Struct. Biol. 7, 399–406.
[23] Clauss, I.M., Wathelet, M.G., Szpirer, J., Islam, M.Q., Levan, G., Szpirer, C. and Huez, G.A. (1991) Human thymosin-beta 4/6–26 gene is part of a multigene family composed of seven members located on seven different chromosomes. Genomics 9, 174–180.
[24] Dhaese, S., Vandepoele, K., Waterschoot, D., Vanloo, B., Vandekerckhove, J., Ampe, C. and Van Troys, M. (2009) The Mouse thymosin beta15 gene family displays unique complexity and encodes a functional thymosin repeat. J. Mol. Biol. 387, 809–825.
[25] Lahn, B.T. and Page, D.C. (1997) Functional coherence of the human Y chromosome. Science 278, 675–680.
[26] Eadie, J.S., Kim, S.W., Allen, P.G., Hutchinson, L.M., Kantor, J.D. and Zetter, B.R. (2000) C-terminal variations in β-thymosin family members specify functional differences in actin-binding properties. J. Cell. Biochem. 77, 277–287.
[27] Huff, T., Zerzawy, D. and Hannappel, E. (1995) Interactions of beta-thymosins, thymosin beta 4-sulfoxide, and N-terminally truncated thymosin beta 4 with actin studied by equilibrium centrifugation, chemical cross-linking and viscometry. Eur. J. Biochem. 230, 650–657.